Received: 19 August 2021 | Revised: 26 October 2021 | Accepted: 27 October 2021

DOI: 10.1111/jsr.13518

Journal of Sleep Research

RESEARCH ARTICLE

Comparing objective wakefulness and vigilance tests to on-the-road driving performance in narcolepsy and idiopathic hypersomnia

Denise Bijlenga[1,2] | Bram Urbanus[1] | Nick N. J. J. M. van der Sluiszen[3] | Sebastiaan Overeem[4,5] | Jan G. Ramaekers[3] | Annemiek Vermeeren[3] | Gert Jan Lammers[1,2]

[1]Sleep-Wake Centre, Stichting Epilepsie Instellingen Nederland (SEIN), Heemstede, the Netherlands

[2]Department of Neurology, Leiden University Medical Centre, Leiden, the Netherlands

[3]Department of Neuropsychology and Psychopharmacology, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, the Netherlands

[4]Centre for Sleep Medicine, Kempenhaeghe, Heeze, the Netherlands

[5]Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, the Netherlands

Correspondence
Denise Bijlenga, Sleep-Wake Centre, Stichting Epilepsie Instellingen Nederland, (SEIN), Achterweg 5, 2103 SW Heemstede, the Netherlands.
Email: dbijlenga@sein.nl

Funding information
Dutch Ministry of Infrastructure and Water Management

Summary

Patients with narcolepsy or idiopathic hypersomnia (IH) are at increased risk of driving accidents. Both excessive daytime sleepiness, i.e. unwanted sleep episodes during the day, and disturbed vigilance are core features of these disorders. We tested on-the-road driving performance of patients with narcolepsy or IH coming in for a routine driving fitness evaluation and examined: (1) correlations between driving performance and the Maintenance of Wakefulness Test (MWT), Sustained Attention to Response Task (SART) and Psychomotor Vigilance Test (PVT) as objective tests; (2) the predictive power of the MWT and SART for increased risk of impaired driving; (3) the best set of objective predictors for increased risk of impaired driving. Participants were 44 patients (aged 18–75 years) with narcolepsy type 1 (NT1), type 2 (NT2) or IH. They completed the MWT, SART, PVT, a subjective sleepiness questionnaire, and a standardised on-the-road driving test. The standard deviation of the lateral position (SDLP) was used as outcome measure of driving performance. The MWT had low correlation with the SDLP ($\rho = -0.41$ to -0.49, $p < 0.01$). The SART and PVT had low correlations with SDLP ($\rho = 0.30$ and $\rho = 0.39$, respectively, both $p < 0.05$). The predictive power of MWT for an increased risk of impaired driving was significant, but low (area under the curve = 0.273, $p = 0.012$), and non-significant for SART. We conclude that in our present group, none of the tests had adequate ability to predict impaired driving, questioning their use for clinical driving fitness evaluation in narcolepsy and IH. Real-time monitoring of sleepiness while driving seems more promising in these patients.

KEYWORDS
fitness to drive, Hypersomnolence, sleepiness, traffic injury prevention

1 | INTRODUCTION

Sleepiness behind the wheel is one of the major causes of driving accidents (Connor et al., 2001; Philip et al., 2010). Patients with the sleep disorder narcolepsy or idiopathic hypersomnia (IH) are at more than twice the risk of being involved in a driving accident, especially in the untreated condition (Philip et al., 2010; Powell et al., 2007). Our previous study showed that also treated patients with narcolepsy type 1 (NT1) may still have an increased risk of impaired driving as compared to controls (van der Sluiszen et al., 2021). Therefore, there is a need for objective tests that can reliably predict increased risk of impaired driving.

Excessive daytime sleepiness (EDS) is a core feature of narcolepsy and IH and is described as the inability to stay awake. It includes both subjective feelings of sleepiness and actual falling asleep during monotonous or potentially dangerous situations. As patients do not always adequately perceive their sleepiness symptoms and may underestimate their degree of EDS, the use of valid objective tests is indicated. Another core deficit in narcolepsy and IH, related to EDS, is a disturbed vigilance (Van Schie et al., 2012; Thomann et al., 2014). Vigilance problems have a major impact on the daily life of these persons, for instance when driving a car (Fronczek et al., 2006; van der Heide, Donjacour, et al., 2015).

Standardised on-the-road driving tests have been used to measure driving fitness in e.g. patient groups and for the evaluation of the effect of various psychoactive substances on driving performance (O'Hanlon, 1984; Ramaekers, 2017). Such driving tests are regarded as the best proxy of actual driving, with the standard deviation of the lateral position (SDLP) as the main outcome. As such standardised driving tests and driving simulators are expensive and not feasible for the clinical evaluation of driving fitness in patients with EDS, we aimed to examine other objective tests that are more feasible and may reliably predict increased risk of impaired driving. Three widely used objective tests in EDS and vigilance involve the Maintenance of Wakefulness Test (MWT), the Sustained Attention to Response Task (SART), and the Psychomotor Vigilance Test (PVT).

The MWT is an objective measurement that assesses one's ability to remain awake, consisting of four times 40 min of rest, with simultaneous recording of brain activity (Littner et al., 2005). It is therefore a time-consuming and costly test. The MWT has been adopted in several European countries in fitness to drive legislations in case of EDS. However, studies evaluating the relationship between the MWT and driving performance in patients with EDS show mixed results. In patients with obstructive sleep apnea correlations varying from low (0.34) to high (0.85) were found between MWT and simulated or on-the-road driving performances (Philip et al., 2008; Pizza et al., 2009; Sagaspe et al., 2007). In narcolepsy, correlations ranging between low (0.26) and moderate (0.56) were found (Philip et al., 2014; Sagaspe et al., 2019). In a recent study among patients with various sleep-related disorders, retrospective MWT outcomes were associated with self-reported near misses or accidents in the past year (Philip et al., 2020). However, these MWT tests were performed for clinical purposes and not for the evaluation of fitness to drive. This is a huge difference, as motivation to stay awake during the MWT, e.g. because one's driving license may be revoked after a failed MWT, may impact the MWT outcomes. This has already been shown in healthy subjects (Bonnet & Arand, 2005). Moreover, a study showed that sleepiness during the MWT cannot be used to judge sleepiness perception while driving (Schreier et al., 2015). Even though the MWT is currently the best available test to objectify one's ability to stay awake, its application for the evaluation of fitness to drive is still debatable (Wise, 2006).

Objective tests focussing on vigilance instead of the ability to stay awake or sleepiness seem to be a more rational approach. Therefore, their potential to assess fitness to drive should be assessed and compared to the MWT, particularly facing the apparently insufficient properties of the MWT in assessing fitness to drive in central disorders of hypersomnolence. One such vigilance tests is the SART, which is a simple computer task of <5 min. The SART is elaborately tested in individuals with central disorders of hypersomnolence and is used as a measure of treatment efficacy in narcolepsy (Dauvilliers et al., 2013; van der Heide, Donjacour, et al., 2015). The SART was originally developed and tested in patients with traumatic brain injury (Robertson et al., 1997) and has shown to be able to distinguish patients with hypersomnolence from healthy controls (Fronczek et al., 2006; Van Schie et al., 2012).

Another widely used vigilance test is the PVT, which is a simple 10-min reaction-time test. The PVT outcomes in patients with narcolepsy and hypersomnia has been shown to be worse as compared to controls (Thomann et al., 2014). The PVT also correlates well with decrements of on-the-road driving performance in sleep-deprived individuals (Jongen et al., 2017).

To date, these objective tasks have not been compared head-to-head in the context of actual driving performance in patients with narcolepsy and IH. In the present study, aimed to examine in patients with narcolepsy or IH: (1) correlations between the SDLP as a measure of driving performance, and the MWT, SART and PVT; (2) the predictive power of the MWT and SART as objective tests for increased risk of impaired driving performance; and (3) the best set of objective tests to predict increased risk of impaired driving performance. Results of this paper will indicate which of these objective tests or set of tests could best be used in the clinical assessment of fitness to drive in patients with narcolepsy and IH.

2 | METHODS

2.1 | Subjects

Participants were consecutive individuals referred for evaluation of their fitness to drive between June 2015 and January 2017 at two Dutch specialised outpatient centres for sleep-wake disorders Kempenhaeghe and SEIN; at the time the only centres evaluating fitness to drive in narcolepsy and IH in the Netherlands. Other inclusion criteria were aged between 18 and 75 years, a diagnosis of (NT1, narcolepsy type 2 (NT2) or IH according to the International Classification of Sleep Disorders, third edition (ICSD-3), and either not using medication

for their hypersomnolence disorder, or having had no changes in the treatment of their hypersomnolence disorder for ≥6 weeks prior to the study. At the time of this study, a mean MWT sleep onset latency (SOL) of >8 min, a score on the self-reported Epworth Sleepiness Scale (ESS) of <11, and the clinical recommendations of the sleep physician, determined the eligibility of drivers with narcolepsy and IH to retain or renew their driving license in the Netherlands and therefore served as another selection criterion in this study. As the subjects were aware that the results of the MWT and ESS could affect their eligibility to retain their driving license, they were motivated to perform well on the MWT. As we deem that the score on the ESS can be easily manipulated in order to gain a better outcome, we did not include the ESS as one of the outcome measures in this study.

The Medical Ethics Committee of Maastricht University and academic hospital Maastricht approved the study (www.toetsingonline.nl, NL50579.068.14). The study was conducted in agreement with the code of ethics on human experimentation established by the Declaration of Helsinki (1964) and subsequent amendments. Participants were instructed verbally and in writing. Signed informed consent was obtained from each participant before enrolment. Participants received a waiver for the standard costs of the fitness to drive evaluations, an incentive of €50 (euros), plus travelling expenses.

2.2 | Design

The study consisted of two test days, 1-week apart. Included subjects all completed the assessments on test day one, comprising of the MWT and SART. See Figure 1 for the timelines of the test days. During the test days, participants were not restrained from their normal habits (e.g. drinking coffee or other stimulating beverages, napping, medication use), aiming to mimic performance in everyday live. On test day 1, the SART was administered 15 min before each of the four MWT trials. Before the first trial, a practice trial of the SART was performed to familiarise with the task and to minimise the consequence of learning effects (van Schie et al., 2014). On test day 2, the participants first completed another SART trial, had a driving practice and completed the subjective Karolinska Sleepiness Scale (KSS) as a measure of subjective sleepiness. Then, they had the 1-hr on-the-road driving test. The highway driving test was only conducted when weather conditions were expected to have minimal influence on the outcomes. After the driving test, participants completed another SART, and a single trial of the psychomotor vigilance task (PVT).

2.3 | Outcome measures

2.3.1 | Maintenance of Wakefulness Test (MWT)

The MWT (Mitler et al., 1982) quantifies the capability to stay awake in a monotonous situation. It consists of four 40-min periods in which individuals are asked to remain awake while sitting comfortably in a quiet, semi-dark room. Appearance of sleep is constantly monitored by electroencephalography. An experienced sleep technician performed the MWTs. The primary outcome measure of the MWT is the mean of four SOLs (in minutes). The SOL is defined as the first occurrence of three consecutive 30-s epochs of Stage 1, or any single 30-s epoch of another sleep stage (Stage 2, Stage 3, or rapid eye movement sleep). This is the '3-epoch' definition. Alternatively, the SOL is defined as the occurrence of 15 s of sleep in one 30-s epoch (the '1-epoch' definition). We added the latter definition to the primary outcome to obtain a more normal range of scores and less of a ceiling effect, which has been found in the 3-epoch definition in controls (Doghramji et al., 1997; Littner et al., 2005). Also, the 1-epoch definition may better identify the appearance of short periods of drowsiness, which is potentially dangerous while driving a car. Patients who did not fell asleep during one of the MWT trials were assigned a mean SOL of 40 min. The MWT latencies were also classified as a short sleep latency group (0–19 min), an intermediate group (20–33 min) and a normal sleep latency group (34–40 min) (Doghramji et al., 1997).

2.3.2 | Sustained Attention to Response Task (SART)

The SART (Robertson et al., 1997) is a computer task that measures speed and accuracy of simple responses to a set of targets and the

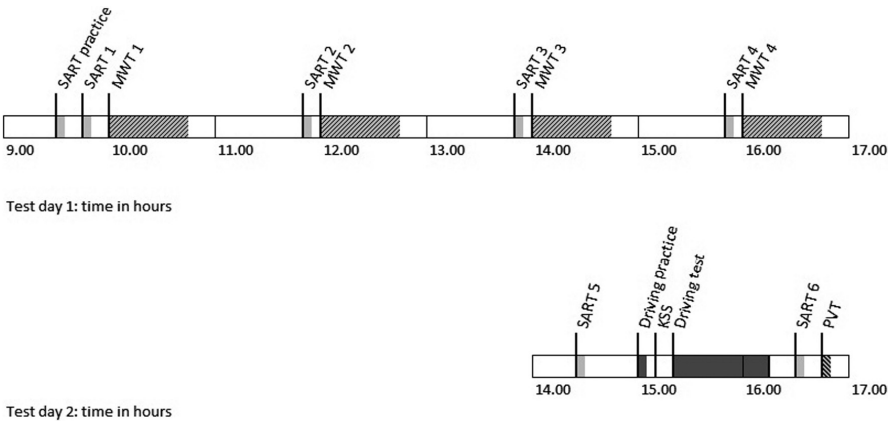

FIGURE 1 Timelines of test days 1 (upper timeline) and 2 (lower timeline). KSS, Karolinska Sleepiness Scale; MWT, Maintenance of Wakefulness Test; PVT, Psychomotor Vigilance Task; SART, Sustained Attention To Response Task

ability to ignore non-targets. In short, it involves pressing a key when a number (1–9) appears on the screen (target), except when the number is a 3 (non-target). Subjects were instructed to respond as accurately as possible (van Schie et al., 2014). The SART contains 225 stimuli and takes 4 min and 20 s. The primary outcome measure of the SART is the error rate, consisting of the number of false responses to non-target (commission errors), and the number of non-responses after a target (omission errors).

2.3.3 | Psychomotor Vigilance Task (PVT)

The PVT is a computer task and is based on a simple visual reaction time test (Dinges & Powell, 1982). The test measures the ability to sustain attention over a period of ~10 min. Participants were required to respond to a visual stimulus on a computer screen by pressing a button with the dominant hand. The visual stimulus is the presentation of a counter that starts running from 0 to 60 s at 1-ms intervals. The participants are required to respond to this visual counter as soon as they perceive it on screen by pressing the corresponding button. If a response is made, the counter stops, stays on screen for 500 ms as visual feedback for the volunteer, and disappears. The inter-stimuli interval is variable (2–10 s). In total, 100 stimuli are presented on the screen. If a response has not been made within 60 s, this is a non-response to a target. Primary outcome measures are mean inverse reaction time (1/RT) and number of lapses (responses with RT of ≥500 ms) (Basner & Dinges, 2011). Performance on the PVT has been calibrated for dose effects of alcohol and 1 night of sleep deprivation (Jongen et al., 2014, 2015).

2.3.4 | On-the-road highway driving test

The on-the-road driving test measures the driving performance in a specially equipped vehicle over a 100 km primary highway circuit and was initially developed to assess drug effects on driving performance (O'Hanlon, 1984; Ramaekers, 2017). Participants were accompanied by a licensed driving instructor having access to dual controls (brakes and accelerator). The participants were instructed to drive with a steady lateral position between the delineated boundaries of the slower (right) traffic lane, while maintaining a constant speed of 95 km/hr (58 mph). They were allowed to deviate from these instructions only to pass a slower vehicle, and to leave and re-enter the highway at the mid-circuit turnaround point. Whenever participants had doubts about their competence to drive safely, they were instructed to terminate the test. If the driving instructor judges their driving to become unsafe, he ordered the participant to stop the vehicle. The main end-point was the mean SDLP (in centimetres) over the entire test, which is a measure of road tracking error, or "weaving". The SDLP has a high test–retest reliability of $r = 0.80$ (Verster & Roth, 2011). The SDLP scores of prematurely terminated tests were calculated from the data collected until termination of each ride. The standard operating procedure for this test is described in (Verster & Roth, 2011). The calculation of the SDLP is described in (van der Sluiszen et al., 2021).

Most previous studies involving SDLP were designed to evaluate the influence of drugs or alcohol on driving performance, using a threshold of a 2.5-cm increase in SDLP as an indication of clinically relevant driving impairment. This threshold is equivalent to the effects of blood alcohol concentrations of 0.5 mg/ml on SDLP, the legal limit for driving in most countries (Jongen et al., 2017). In our previous study, the mean difference in the SDLP was 1.02 cm between our hypersomnolence patients ($n = 45$, mean [*SE*] 18.68 [0.56] cm) and healthy controls ($n = 31$, mean [*SE*] 17.66 [0.67] cm), which was non-significant and lower than this cut-off. Interestingly, the 95% confidence interval (CI) of the mean difference of the SDLP between the hypersomnolence and control group included the non-inferiority limit (i.e. +2.5 cm) and zero and was therefore considered inconclusive. This indicates more inter-individual differences in driving performance between patients with hypersomnolence (van der Sluiszen et al., 2021). The healthy control group ($n = 76$) of another of our previous study, including the same methodology, had a SDLP with a mean (*SE*) of 18.19 (0.46) cm (Vinckenbosch et al., 2021). In the current study, the cut-off of the SDLP to indicate increased risk of impaired driving was set at 19.09 cm, which was the upper limit of the two-sided 95% CI of the *SE* of that previous study. Subjects who were ordered by the driving instructor to terminate the driving test due to risky driving were also classified as increased risk of impaired driving. Subjects with SDLPs below the cut-off who terminated the driving tests themselves were not considered to have increased risk of impaired driving, as they acted correctly upon their self-perceived sleepiness by stopping the vehicle.

2.3.5 | Karolinska Sleepiness Scale (KSS)

In order to get a sense of the momentary subjective feeling of sleepiness before the driving test, the participants completed the KSS, involving 9-point scales ranging from 1 ("very alert") to 9 ("very sleepy, fighting sleep, an effort to keep awake") (Akerstedt & Gillberg, 1990).

2.4 | Analyses

Patients' characteristics were displayed using means and standard deviations (*SD*s) for continuous data and with frequencies for categorical data, for the total group, the driving group, and for the group that was unfit to drive. As data of SART errors and MWT SOLs were not normally distributed, non-parametric analyses were used.

The Mann–Whitney *U* test was used to test for differences in test outcomes between those with increased risk of impaired driving and those with normal driving performance. The chi-square test was used to examine differences in categorical distributions between groups. Spearman correlations (ρ) were used to examine associations between test outcomes of the MWT, SART, PVT, KSS and SDLP. Correlations of ≤0.5 were regarded as low, between 0.5 and 0.7 moderate, and ≥0.7 high (Hinkle et al., 2003).

To assess the predictive value of the MWT and SART on test day 1 on increased risk of impaired driving on test day 2, we applied receiver operating characteristics (ROC) curves. The area under the

curve (AUC), sensitivity and specificity were calculated. In order to define the optimal set of tests to predict increased risk of impaired driving, we applied a binary logistic regression with backward elimination based on least residuals, with age, MWT, and SART of test day 1 as continuous, and gender as binary, predictors variables.

An α-level of 0.05 (two-tailed) was used to indicate statistical significance. All statistical analyses were performed using the IBM Statistical Package for the Social Sciences (SPSS®), version 24.0 (IBM Corp., Armonk, NY, USA).

3 | RESULTS

3.1 | Subjects

Participant characteristics are presented in Table 1. There were 88 eligible patients, of whom 45 participated in the on-the-road driving test. Due to a technical error, driving data of one participant were not recorded. Of the 44 participants included in the analyses, 36% were female and the mean (SD, range) age was 42.1 (15.8, 18–74) years. Of the participants, 31 (71%) had NT1, seven (16%) had NT2 and six (14%) had IH, of whom 41 (93%) used medication for their hypersomnolence disorder: 75% used stimulants, 43% used sodium oxybate, and 25% used both. None of the patient characteristics were significantly different between the initial group of eligible patients and those participating in the driving test. The SDLP score was >19.09 cm in 17 (39%) of the participants and they were therefore classified as at increased risk of impaired driving. Two participants (both NT1) were classified as at increased risk of impaired driving because the driving test was terminated by the driving instructor; these participants also had an SDLP of >19.09 cm. Four participants (all NT1) terminated the driving test prematurely due to self-observed experienced sleepiness. Other patient characteristics are shown in Table 1. None of the patient characteristics were significantly different between the total driving group and those with increased risk of impaired driving.

3.2 | Outcome measures

The MWT 3-epoch and 1-epoch outcomes of the four MWT trials on day 1 did not differ between trials (all comparisons $p > 0.05$). Of the participants, 30% did not fall asleep during the MWT trials using the 3-epoch definition and 25% using the 1-epoch definition, obtaining the maximum score of 40 min. Between diagnoses, there were differences in mean SOLs ($F(2,41) = 5.46$, $p = 0.008$); Tukey's post hoc tests showed that patients with IH had significantly longer SOLs than NT1 patients (MWT 3-epoch definition: mean [*SD*] 39.2 [2.0] versus 28.1 [9.8] min, $p = 0.016$; MWT 1-epoch definition: 38.0 [3.2] versus 24.6 [10.6] min, $p = 0.008$).

The SART error scores were not significantly different between the four trials on day 1 (all comparisons $p > 0.05$). The SART total error scores did not differ between diagnostic groups ($F(2,41) = 0.830$, $p = 0.443$), and the SART scores of day 1 and day 2 did not differ significantly ($t(43) = 0.693$, $p = 0.492$). As the within-test outcomes of the separate trials did not differ, we used the individual summary scores based on means over all trials for the MWT and SART. Table 2 shows outcomes on the MWT, SART, PVT, and KSS of the total group and also divided into the normal and increased risk of impaired driving groups.

There was a significant difference between the group with normal and increased risk of impaired driving for the MWT 1-epoch definition ($U = 125.50$, $Z = -2.528$, $p = 0.011$). This was also reflected in the distribution over the MWT categories: those with increased risk of impaired driving had less often normal MWT scores (34–40 min) and more often short MWT latencies (0–19 min) than those in the normal driving groups ($\chi^2(2) = 6.25$, $p = 0.044$). Other tests outcomes were not significantly different between those with normal driving and increased risk of impaired driving.

3.3 | Correlations between tests

Table 3 shows a correlation matrix for the test outcomes and driving performance. Within tests, there was a high correlation between the MWT 1-epoch and 3-epoch definitions ($\rho = 0.941$); a moderate correlation between the SART error counts of day 1 and 2 ($\rho = 0.690$), and between the PVT outcomes 1/RT and number of lapses ($\rho = -0.692$).

Between tests, there were low but significant correlations between the MWT 1-epoch definition and PVT lapses ($\rho = -0.316$), and between the SART error counts and PVT 1/RT and number of lapses if performed on the same day ($\rho = -0.371$ and $\rho = 0.304$, respectively). Focussing on only the SART after the driving test (i.e. the second SART on day 2), the correlation with the PVT outcomes are comparable (for PVT number of lapses: $\rho = 0.311$, $p = 0.043$; for PVT 1/RTs: $\rho = -0.311$, $p = 0.042$; not shown in Table 3).

Tests that correlated with SDLP were all low correlations: MWT 3-epoch definition ($\rho = -0.405$), MWT 1-epoch definition ($\rho = -0.491$, Figure 2A), SART of day 2 ($\rho = 0.300$, Figure 2B), PVT 1/RTs ($\rho = -0.380$), and PVT number of lapses ($\rho = 0.389$, Figure 2C).

Focussing on only the SART after the driving test (i.e. the second SART on day 2), the correlation with SDLP was non-significant ($\rho = -0.276$, $p = 0.070$; not shown in Table 3).

3.4 | Predictive value of MWT and SART for increased risk of impaired driving

The ability of the 1-epoch definition of the MWT to predict increased risk of impaired driving was significant, although low (AUC = 0.273, $p = 0.012$). An MWT of ≤19 min had a sensitivity (correctly rated as increased risk of impaired driving) of 53% and a specificity (incorrectly rated as normal driving) of 85%. An optimal cut-off for sensitivity and specificity could not be reached. The MWT 3-epoch definition and the SART were both not able to significantly predict increased risk of impaired driving (MWT 3-epoch: AUC = 0.330, $p = 0.087$; SART: AUC = 0.564, $p = 0.477$).

TABLE 1 Patient characteristics of the eligible participants (n = 88), those of whom participated in the driving test (total driving group; N = 44), and those of whom had increased risk of impaired driving (n = 17)

Patient characteristics	Eligible patients $n = 88$	Total driving group $N = 44$	Increased risk of impaired driving $n = 17$
Age, years, mean (*SD*, range)	38.3 (13.7, 18.1–74.8)	42.1 (15.8, 18.1–74.8)	40.7 (17.1, 18.1–67.3)
Females, *n* (%)	37 (42.0)	16 (36.4)	3 (17.6)
Diagnoses, *n* (%)			
Narcolepsy type 1	54 (61.4)	31 (70.5)	14 (82.0)
Narcolepsy type 2	15 (17.0)	7 (15.9)	0
Idiopathic hypersomnia	19 (21.6)	6 (13.6)	3 (17.6)
Medication, *n* (%)			
Stimulants	67 (76.1)	33 (75.0)	12 (70.6)
Sodium oxybate (Xyrem ®)	29 (33.0)	19 (43.2)	7 (41.2)
Both stimulants and sodium oxybate	18 (20.4)	11 (25.0)	2 (11.8)
Antidepressants	10 (11.4)	5 (11.4)	3 (17.6)
No medication	7 (8.0)	3 (6.8)	2 (11.8)
Driving license, years, mean (*SD*, range)	–	19.9 (14.6, 0–53)	21.3 (17.5, 0–50)
Driving experience, km/year, mean (*SD*, range)	–	9.9k ± 10.5k [0–35k]	8.7k ± 9.8k [0-30k]
Fitness to drive examination, *n* (%)			
First examination	42 (47.7)	20 (45.5)	7 (41.2)
Second examination	25 (28.4)	12 (27.3)	5 (29.4)
Third examination	21 (23.9)	12 (27.3)	5 (29.4)
Caffeine, ≥5 units before the driving test, *n* (%)	–	4 (9.2)	1 (5.9)
Alcohol, ≥2 units/day, *n* (%)	–	9 (20.4)	2 (11.8)
Smoking, yes, *n* (%)	–	15 (34.1)	8 (47.1)
Driving test terminated by participant, *n* (%)	–	4 (9.0)	2 (11.8)
Driving test SDLP, cm, mean (*SD*)	–	18.7 (4.0)	22.7 (3.0)

[Correction added on 22 December 2021, after first online publication: The MWT definition in the preceding sentence has been corrected to 3-epoch.]

3.5 | Predictive value of a combination of parameters

The backward elimination binary logistic regression showed that after five steps, the MWT 1-epoch definition alone could correctly classify 66% of cases into impaired or normal driving, with low explained variance ($R^2 = 0.20$, $p = 0.008$; Table 4). No set of tests was identified that could reliably predict increased risk of impaired driving.

4 | DISCUSSION

The present study investigated driving performance in patients with narcolepsy and IH coming in for a routine fitness to drive evaluation, and examined correlations between driving performance and outcomes on the objective MWT, SART, and PVT. We examined the MWT and SART as possible objective predictors of driving performance in these patients. Results showed that: (1) correlations between driving performance and MWT, SART or PVT were low at best; (2) the MWT is insufficiently, and the SART is not, able to predict increased risk of impaired driving, as defined by an SDLP score of >19.09 cm; and (3) no set of tests could be identified to reliably predict risk of impaired driving in mostly treated patients with narcolepsy or IH during a fitness to drive evaluation.

Our present patients were either on stable treatment for narcolepsy and IH or did not use medication. Also, they were very motivated to stay awake as their driver's license may be revoked in case of poor MWT results. They exhibited a sleep latency on the MWT that is similar to results found in healthy subjects (Doghramji et al., 1997). The same holds true for the SART, as >50% of patients performed comparable to healthy controls (Fronczek et al., 2006) and error levels were clearly lower than in untreated hypersomnolence (Van Schie et al.,

TABLE 2 Means and *SD*s on the MWT, SART, PVT, KSS, driving performances, and the distribution into MWT categories of the total group ($N = 44$), the normal driving performance group ($n = 27$), and increased risk of impaired driving performance group ($n = 17$), and the comparison between the normal and increased risk of impaired driving groups

	Total group $N = 44$	Normal driving $n = 27$	Increased risk of impaired driving $n = 17$	p Normal versus increased risk of impaired driving
Day 1 test outcomes				
MWT (3-ep) SOL, min, mean (*SD*)	30.8 (9.4)	33.0 (8.0)	27.2 (10.6)	0.057
MWT (1-ep) SOL, min, mean (*SD*)	27.6 (10.5)	30.9 (9.4)	22.5 (10.3)	0.011*
MWT (1-ep) SOL categories, *n* (%)				
Short: 0–19 min	12 (27.3)	4 (14.8)	8 (47.1)	0.044*
Intermediate: 20–33 min	17 (38.6)	11 (40.7)	6 (35.3)	
Normal: 34–40 min	15 (34.1)	12 (44.4)	3 (17.6)	
SART total error count, mean (*SD*)	4.4 (3.7)	3.9 (3.2)	5.1 (4.5)	0.476
Day 2 test outcomes, mean (*SD*)				
SART total error count	4.0 (4.2)	3.4 (3.8)	5.0 (4.7)	0.195
PVT 1/RT, s	3.8 (0.4)	3.9 (0.4)	3.7 (0.4)	0.379
PVT lapses (number)	1.1 (2.3)	0.5 (1.1)	2.0 (3.3)	0.137
KSS before driving	3.1 (1.5)	3.3 (1.9)	2.8 (0.7)	0.887

ep, epoch; KSS, Karolinska Sleepiness Scale; MWT, Maintenance of Wakefulness Test; PVT, Psychomotor Vigilance Task; 1/RT, inverse reaction time; SART, Sustained Attention To Response Task; SOL, sleep onset latency.

*$p < 0.05$.

TABLE 3 Spearman correlations matrix of the outcome measures and between driving performance and test performance, $N = 44$

	Day 1 test outcomes			Day 2 test outcomes			
	MWT (3-ep)	MWT (1-ep)	SART total	SART total	PVT 1/RT	PVT lapses	KSS before
Day 1 test outcomes							
MWT (1-ep) SOL (min)	0.941**	–	–	–	–	–	–
SART total error count	−0.098	−0.144	–	–	–	–	–
Day 2 test outcomes							
SART total error count	−0.131	−0.138	0.690**	–	–	–	–
PVT 1/RT (s)	0.163	0.170	−0.277	−0.371*	–	–	–
PVT lapses (*n*)	−0.245	−0.316*	0.133	0.304*	−0.692**	–	–
KSS before driving	−0.173	−0.132	0.031	−0.102	−0.114	−0.177	–
Day 2 driving performance							
Driving test SDLP	−0.405**	−0.491**	0.211	0.300*	−0.380*	0.389**	0.055

ep, epoch; KSS, Karolinska Sleepiness Scale; MWT, Maintenance of Wakefulness Test; PVT, Psychomotor Vigilance Task; 1/RT, inverse reaction time; SART, Sustained Attention To Response Task; SDLP, standard deviation of lateral position; SOL, sleep onset latency.

*$p < 0.05$.; **$p < 0.01$.

2012). These findings therefore suggest that stable treatment and high intrinsic motivation have beneficial effects on different components of vigilance and test results, as also shown previously (Philip et al., 2014). The patients with NT1 were objectively sleepier than those with IH, given lower MWT sleep latencies. However, vigilance measured with the SART was not different across diagnosis types, as also reported previously (Van Schie et al., 2012).

As the MWT is an elaborate and expensive test, shorter and cheaper alternatives for the assessment of driving fitness should be taken into account. The MWT was the only test that had significantly different outcomes between the normal driving group and the group at increased risk of impaired driving. Also, the MWT had the best predictive value for driving performance, although its predictive power is too low for its application in clinical practice for driving fitness evaluation. Previous studies showed better correspondence

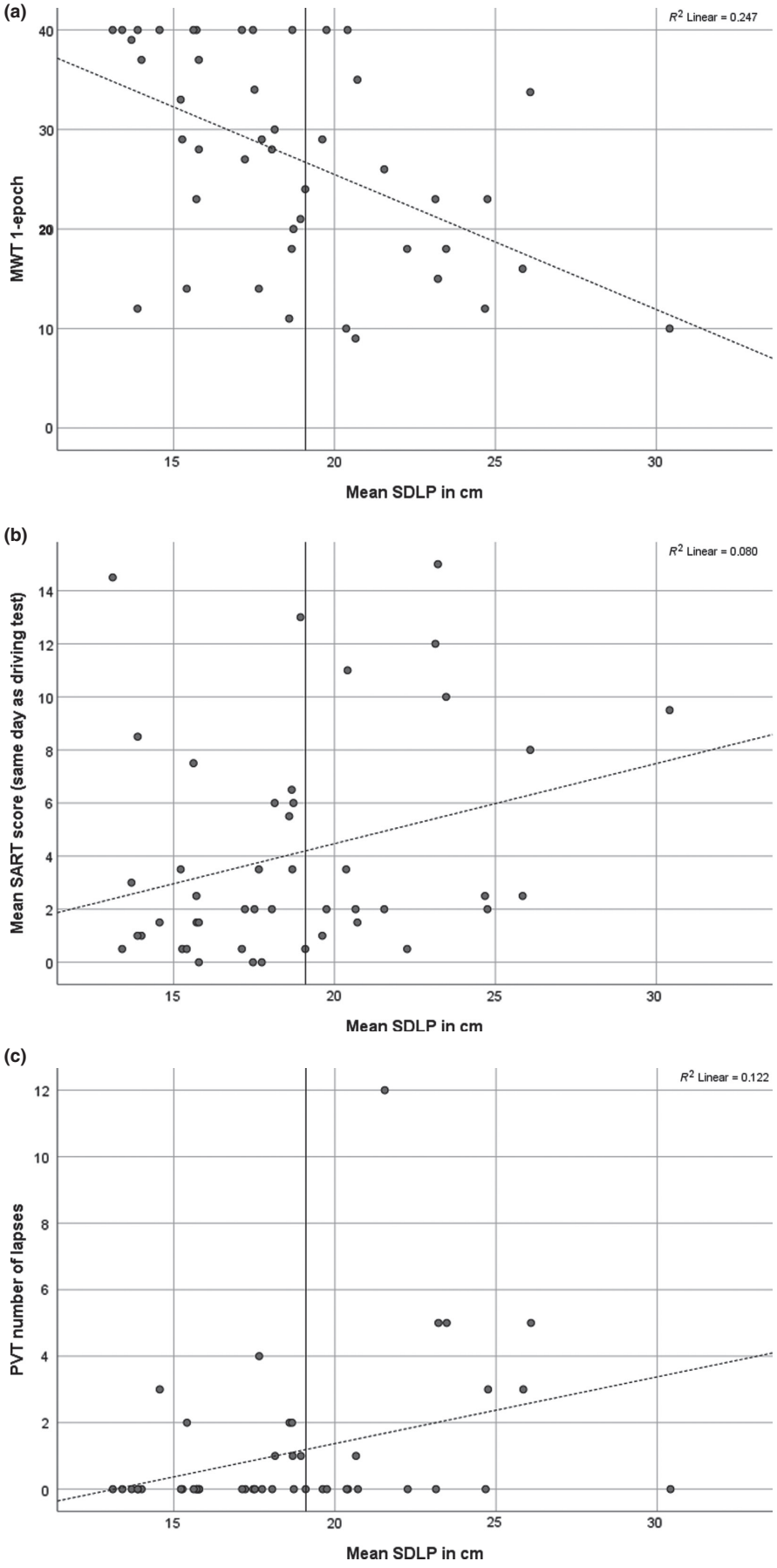

FIGURE 2 Scatterplot of the mean sleep onset latency (SOL) on the Maintenance of Wakefulness Test (MWT) using the 1-epoch definition and the mean standard deviation of the lateral position (SDLP) of the on-the road driving test, $N = 44$, with (a) the mean SOL on the MWT using the 1-epoch definition; (b) the Sustained Attention To Response Task (SART) sum score on the same day as the driving test; (c) the number of lapses on the Psychomotor Vigilance Task (PVT). The vertical lines represent the cut-off for increased risk of impaired driving (SDLP = 19.09 cm); the dashed lines represent linear interpolation

TABLE 4 Step-backward binary logistic regression for combination of tests and demographic variables to predict increased risk for impaired driving, $N = 44$

Step	Predictor variables	Nagelkerke R^2	% correct	Sig. p	Sig. change of step
1	Age, gender, MWT 3-ep, MWT 1-ep, SART	0.270	72.7	0.082	0.082
2	Gender, MWT 3-ep, MWT 1-ep, SART	0.269	72.7	0.045	0.840
3	Gender, MWT 3-ep, MWT 1-ep	0.264	75.0	0.023	0.651
4	Gender, MWT 1-ep	0.246	75.0	0.012	0.392
5	MWT 1-ep	0.201	65.9	0.008	0.189

ep, epoch; MWT, Maintenance of Wakefulness Test; SART, Sustained Attention To Response Task.

between the MWT and driving in mixed groups of patients with EDS (Philip et al., 2013). In a recent study, the SDLP of a driving simulator had a low correlation of $r = 0.34$ with the on-the-road driving outcomes in a similar group of patients, showing that a driving simulator does not outperform the MWT (Sagaspe et al., 2019). That study had a comparable correlation of $r = -0.56$ between the MWT 1-epoch definition and the SDLP of the on-the-road driving test, as in our present study ($\rho = -0.49$).

A shorter and cheaper alternative to the MWT to evaluate driving fitness was not found in the present study. The SART and PVT both had low correlations with SDLP. SART error scores had no correlation with MWT sleep latency, confirming the results from a study comparing the SART and MWT in evaluating the effects of treatment (van der Heide, van Schie, et al., 2015). The EDS as measured by the MWT, and vigilance as measured with the SART, thus comprise totally different constructs. While the SART is useful to assess treatment effects, it is not suitable to predict driving performance in narcolepsy and IH. However, the predictive value of the relatively expensive and time consuming MWT is too low and therefore we conclude that the also the MWT is not a suitable test to assess driving fitness in patients with central disorders of hypersomnolence.

Several limitations should be addressed. First of all, we could only include individuals in possession of a valid driver's license, as this required by Dutch legislation. Excluding potentially more severely affected patients (with an MWT of <8 min) may have masked the predictive value of the tests. Also, our main outcome, the SDLP, involves just one aspect of on-the-road driving. Whereas SDLP can be regarded as a measure of overall vehicle control, it does not provide information on the specific skills and abilities that led to performance impairment. Brief but potentially dangerous changes in alertness may not be captured with the SDLP (Hood & Bruck, 1996). However, the SDLP is the most reliable known outcome measure of driving impairment to date (Vinckenbosch et al., 2020). Third, we argue that patients who correctly act upon their sleepiness behind the wheel by stopping the vehicle, are not regarded as a potential risk in traffic. Therefore, we did not regard patients who decided to stop the driving test due to sleepiness as impaired drivers. However, clinicians may judge otherwise if this was reported in a clinical setting. Lastly, to allow analysis of sensitivity and specificity of the tests we used an ad hoc criterion for increased risk of impaired driving. This criterion is intended for research purposes only and cannot be used as a clinical cut-off point.

Treatment of narcolepsy and IH with modafinil or solriamfetol have shown to improve driving performance (Philip et al., 2014; Vinckenbosch et al., Submitted). Other treatments such as with stimulants, sodium oxybate, and lifestyle changes have not been evaluated in regard to driving. A recent review concluded that treated patients with narcolepsy may be able to drive safely with appropriate limitations (McCall & Watson, 2020). The person's ability to judge his or her state of sleepiness and driving safety, and ability to act accordingly, may be the most important factor indicating driving fitness. However, this is a subjective state of mind that is hard to evaluate objectively. Currently, the MWT is used as one of the criteria to evaluate driving fitness in several countries. In the present study, we showed that the predictive power of the WMT was low in our group of mostly medicated patients, and its application is questionable as a measure to predict fitness to drive in (treated) patients with narcolepsy and IH.

In our present study, the MWT, PVT and SART were not reliable objective tests for the clinical evaluation of driving fitness in central disorders of hypersomnolence. Future studies aiming to lower driving risk may need to focus on real-time monitoring solutions such as detection of sleepiness before and during driving.

ACKNOWLEDGEMENTS

We thank the Dutch Ministry of Infrastructure and Water Management for their financial support of this study. We thank Bas Schottert for performing the MWT and SART.

CONFLICT OF INTEREST

None of the authors have any conflicts of interest.

AUTHOR CONTRIBUTIONS

DB: data analysis and writing of final report; BU: data collection, data processing and analysis, writing of draft report; NS: data collection, data processing and analysis; SO: study design, data collection; JR: study design; AV: study design, data collection; GL: study design, data collection, writing of final report. All authors have read and approved the final report.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Denise Bijlenga https://orcid.org/0000-0002-5169-9830
Nick N. J. J. M. van der Sluiszen https://orcid.org/0000-0002-4523-4519

REFERENCES

Akerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, *52*(1–2), 29–37. https://doi.org/10.3109/00207459008994241

Basner, M., & Dinges, D. F. (2011). Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep*, *34*(5), 581–591. https://doi.org/10.1093/sleep/34.5.581

Bonnet, M. H., & Arand, D. L. (2005). Impact of motivation on multiple sleep latency test and maintenance of wakefulness test measurements. *Journal of Clinical Sleep Medicine*, *1*(4), 386–390. https://doi.org/10.5664/jcsm.26367

Connor, J., Whitlock, G., Norton, R., & Jackson, R. (2001). The role of driver sleepiness in car crashes: A systematic review of epidemiological studies. *Accident Analysis and Prevention*, *33*(1), 31–41. https://doi.org/10.1016/s0001-4575(00)00013-0

Dauvilliers, Y., Bassetti, C., Lammers, G. J., Arnulf, I., Mayer, G., Rodenbeck, A., … Schwartz, J.-C. (2013). Pitolisant versus placebo or modafinil in patients with narcolepsy: A double-blind, randomised trial. *The Lancet Neurology*, *12*(11), 1068–1075. https://doi.org/10.1016/S1474-4422(13)70225-4

Dinges, D. F., & Powell, J. W. (1982). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods*, *17*, 652–655.

Doghramji, K., Mitler, M. M., Sangal, R. B., Shapiro, C., Taylor, S., Walsleben, J., … Youakim, J. M. (1997). A normative study of the maintenance of wakefulness test (MWT). *Electroencephalography and Clinical Neurophysiology*, *103*(5), 554–562. https://doi.org/10.1016/s0013-4694(97)00010-2

Fronczek, R., Middelkoop, H. A., van Dijk, J. G., & Lammers, G. J. (2006). Focusing on vigilance instead of sleepiness in the assessment of narcolepsy: High sensitivity of the Sustained Attention to Response Task (SART). *Sleep*, *29*(2), 187–191.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied Statistics for the Behavioral Sciences* (5th ed.). Cengage Learning Inc.

Hood, B., & Bruck, D. (1996). Sleepiness and performance in narcolepsy. *Journal of Sleep Research*, *5*(2), 128–134. https://doi.org/10.1046/j.1365-2869.1996.00018.x

Jongen, S., Perrier, J., Vuurman, E. F., Ramaekers, J. G., & Vermeeren, A. (2015). Sensitivity and validity of psychometric tests for assessing driving impairment: Effects of sleep deprivation. *PLoS One*, *10*(2), e0117045. https://doi.org/10.1371/journal.pone.0117045

Jongen, S., Vermeeren, A., van der Sluiszen, N. N., Schumacher, M. B., Theunissen, E. L., Kuypers, K. P., … Ramaekers, J. G. (2017). A pooled analysis of on-the-road highway driving studies in actual traffic measuring standard deviation of lateral position (i.e., "weaving") while driving at a blood alcohol concentration of 0.5 g/L. *Psychopharmacology (Berl)*, *234*(5), 837–844. https://doi.org/10.1007/s00213-016-4519-z

Jongen, S., Vuurman, E., Ramaekers, J., & Vermeeren, A. (2014). Alcohol calibration of tests measuring skills related to car driving. *Psychopharmacology (Berl)*, *231*(12), 2435–2447. https://doi.org/10.1007/s00213-013-3408-y

Littner, M. R., Kushida, C., Wise, M., G. Davila, D., Morgenthaler, T., Lee-Chiong, T., … & Kramer, M (2005). Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test. *Sleep*, *28*(1), 113–121. https://doi.org/10.1093/sleep/28.1.113

McCall, C. A., & Watson, N. F. (2020). Therapeutic strategies for mitigating driving risk in patients with narcolepsy. *Therapeutics and Clinical Risk Management*, *16*, 1099–1108. https://doi.org/10.2147/TCRM.S244714

Mitler, M. M., Gujavarty, K. S., & Browman, C. P. (1982). Maintenance of wakefulness test: A polysomnographic technique for evaluation treatment efficacy in patients with excessive somnolence. *Electroencephalography and Clinical Neurophysiology*, *53*(6), 658–661. https://doi.org/10.1016/0013-4694(82)90142-0

O'Hanlon, J. F. (1984). Driving performance under the influence of drugs: Rationale for, and application of, a new test. *British Journal of Clinical Pharmacology*, *18*(Suppl 1), 121S–129S. https://doi.org/10.1111/j.1365-2125.1984.tb02590.x

Philip, P., Chaufton, C., Taillard, J., Capelli, A., Coste, O., Léger, D., … Sagaspe, P. (2014). Modafinil improves real driving performance in patients with hypersomnia: A randomized double-blind placebo-controlled crossover clinical trial. *Sleep*, *37*(3), 483–487. https://doi.org/10.5665/sleep.3480

Philip, P., Chaufton, C., Taillard, J., Sagaspe, P., Léger, D., Raimondi, M., … Capelli, A. (2013). Maintenance of Wakefulness Test scores and driving performance in sleep disorder patients and controls. *International Journal of Psychophysiology*, *89*(2), 195–202. https://doi.org/10.1016/j.ijpsycho.2013.05.013

Philip, P., Guichard, K., Strauss, M., Léger, D., Pepin, E., Arnulf, I., … Dauvilliers, Y. (2020). Maintenance of wakefulness test: How does it predict accident risk in patients with sleep disorders? *Sleep Medicine*, *77*, 249–255. https://doi.org/10.1016/j.sleep.2020.04.007

Philip, P., Sagaspe, P., Lagarde, E., Leger, D., Ohayon, M. M., Bioulac, B., … Taillard, J. (2010). Sleep disorders and accidental risk in a large group of regular registered highway drivers. *Sleep Medicine*, *11*(10), 973–979. https://doi.org/10.1016/j.sleep.2010.07.010

Philip, P., Sagaspe, P., Taillard, J., Chaumet, G., Bayon, V., Coste, O., … Guilleminault, C. (2008). Maintenance of Wakefulness Test, obstructive sleep apnea syndrome, and driving risk. *Annals of Neurology*, *64*(4), 410–416. https://doi.org/10.1002/ana.21448

Pizza, F., Contardi, S., Mondini, S., Trentin, L., & Cirignotta, F. (2009). Daytime sleepiness and driving performance in patients with obstructive sleep apnea: Comparison of the MSLT, the MWT, and a simulated driving task. *Sleep*, *32*(3), 382–391. https://doi.org/10.1093/sleep/32.3.382

Powell, N. B., Schechtman, K. B., Riley, R. W., Guilleminault, C., Chiang, R. P., & Weaver, E. M. (2007). Sleepy driver near-misses may predict accident risks. *Sleep*, *30*(3), 331–342. https://doi.org/10.1093/sleep/30.3.331

Ramaekers, J. G. (2017). Drugs and driving research in medicinal drug development. *Trends in Pharmacological Sciences*, *38*(4), 319–321. https://doi.org/10.1016/j.tips.2017.01.006

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). 'Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747–758. https://doi.org/10.1016/S0028-3932(97)00015-8

Sagaspe, P., Micoulaud-Franchi, J.-A., Coste, O., Léger, D., Espié, S., Davenne, D., … Philip, P. (2019). Maintenance of Wakefulness test, real and simulated driving in patients with narcolepsy/hypersomnia. *Sleep Medicine*, *55*, 1–5. https://doi.org/10.1016/j.sleep.2018.02.009

Sagaspe, P., Taillard, J., Chaumet, G., Guilleminault, C., Coste, O., Moore, N., … Philip, P. (2007). Maintenance of wakefulness test as a predictor of driving performance in patients with untreated obstructive sleep apnea. *Sleep*, *30*(3), 327–330.

Schreier, D. R., Roth, C., & Mathis, J. (2015). Subjective perception of sleepiness in a driving simulator is different from that in the

Maintenance of Wakefulness Test. *Sleep Medicine*, *16*(8), 994–998. https://doi.org/10.1016/j.sleep.2015.01.023

Thomann, J., Baumann, C. R., Landolt, H. P., & Werth, E. (2014). Psychomotor vigilance task demonstrates impaired vigilance in disorders with excessive daytime sleepiness. *Journal of Clinical Sleep Medicine*, *10*(9), 1019–1024. https://doi.org/10.5664/jcsm.4042

van der Heide, A., Donjacour, C. E. H. M., Pijl, H., Reijntjes, R. H. A. M., Overeem, S., Lammers, G. J., … Fronczek, R. (2015). The effects of sodium oxybate on core body and skin temperature regulation in narcolepsy. *Journal of Sleep Research*, *24*(5), 566–575. https://doi.org/10.1111/jsr.12303

van der Heide, A., van Schie, M. K. M., Lammers, G. J., Dauvilliers, Y., Arnulf, I., Mayer, G., … van Dijk, J. G. (2015). Comparing treatment effect measurements in narcolepsy: The sustained attention to response task, Epworth sleepiness scale and maintenance of wakefulness test. *Sleep*, *38*(7), 1051–1058. https://doi.org/10.5665/sleep.4810

van der Sluiszen, N., Urbanus, B., Lammers, G. J., Overeem, S., Ramaekers, J. G., & Vermeeren, A. (2021). On-the-road driving performance of patients with central disorders of hypersomnolence. *Traffic Injury Prevention*, *22*(2), 120–126. https://doi.org/10.1080/15389588.2020.1862804

van Schie, M. K., Alblas, E. E., Thijs, R. D., Fronczek, R., Lammers, G. J., & van Dijk, J. G. (2014). The influences of task repetition, napping, time of day, and instruction on the sustained attention to response task. *Journal of Clinical and Experimental Neuropsychology*, *36*(10), 1055–1065. https://doi.org/10.1080/13803395.2014.968099

Van Schie, M. K., Thijs, R. D., Fronczek, R., Middelkoop, H. A., Lammers, G. J., & Van Dijk, J. G. (2012). Sustained attention to response task (SART) shows impaired vigilance in a spectrum of disorders of excessive daytime sleepiness. *Journal of Sleep Research*, *21*(4), 390–395. https://doi.org/10.1111/j.1365-2869.2011.00979.x

Verster, J. C., & Roth, T. (2011). Standard operation procedures for conducting the on-the-road driving test, and measurement of the standard deviation of lateral position (SDLP). *International Journal of General Medicine*, *4*, 359–371. https://doi.org/10.2147/IJGM.S19639

Vinckenbosch, F. R. J., Lammers, G. J., Overeem, S., Chen, D., Wang, G., Carter, L. P., … Vermeeren, A. (Submitted). Effects of solriamfetol on driving in participants with narcolepsy: A randomized crossover trial.

Vinckenbosch, F. R. J., Vermeeren, A., Verster, J. C., Ramaekers, J. G., & Vuurman, E. F. (2020). Validating lane drifts as a predictive measure of drug or sleepiness induced driving impairment. *Psychopharmacology (Berl)*, *237*(3), 877–886. https://doi.org/10.1007/s00213-019-05424-8

Vinckenbosch, F. R. J., Vermeeren, A., Vuurman, E. F. P. M., Sluiszen, N. N. J. J. M., Verster, J. C., Loo, A. J. A. E., Dijken, J. H., Veldstra, J. L., Brookhuis, K. A., De Waard, D., & Ramaekers, J. G. (2021). An explorative approach to understanding individual differences in driving performance and neurocognition in long-term benzodiazepine users. *Human Psychopharmacology*, *36*(4), e2778. https://doi.org/10.1002/hup.2778

Wise, M. S. (2006). Objective measures of sleepiness and wakefulness: application to the real world? *Journal of Clinical Neurophysiology*, *23*(1), 39–49. https://doi.org/10.1097/01.wnp.0000190416.62482.42

How to cite this article: Bijlenga, D., Urbanus, B., van der Sluiszen, N. N. J. J. M., Overeem, S., Ramaekers, J. G., Vermeeren, A., & Lammers, G. J. (2022). Comparing objective wakefulness and vigilance tests to on-the-road driving performance in narcolepsy and idiopathic hypersomnia. *Journal of Sleep Research*, 31, e13518. https://doi.org/10.1111/jsr.13518